Multitask Observation using Satellite Imagery and Kitchen Sinks (MOSAIKS)

Togo Data Lab Training – UCSB, CEGA, & emLab
January, 2025

Tamma Carleton (UC Berkeley & emLab)
in collaboration with: Jonathan Proctor, Trinetta Chong, Taryn Fransen, Simon Greenhill, Jessica Katz, Hikari Murayama, Luke Sherman, Jeanette Tseng, Hannah Druckenmiller, Solomon Hsiang

A growing set of measurements from space

Science

Using satellite imagery to understand and promote sustainable development

Coastal Engineering

Multispectral satellite imagery and machine learning for the extraction of shoreline indicators

Emma McAllister, Andres Payo, Alessandro Novellino, Tony Dolphin, Encarni Medina-Lopez

Computers and Electronics in Agriculture

High-resolution satellite imagery applications in crop phenotyping: An overview

Chongyuan Zhang, Afef Marzougui, Sindhuja Sankaran

Forest Biodiversity Monitoring Based on Remotely Sensed Spectral Diversity—A Review

A growing set of measurements from space

Each measurement or monitoring system is costly to implement in practice.

A growing set of measurements from space

Each measurement or monitoring system is costly to implement in practice.

Source: Farmonaut

For what **outcomes**, in what **places**, and for which **populations** should we invest in satellite-based monitoring?

Evaluating satellite-based predictions for over 100 variables

What can we see from space?

- Existing satellite imagery and machine learning (SIML) predictions are customized and heterogeneous
- This limits our ability to assess what can and cannot be feasibly monitored using this new technology

Evaluating satellite-based predictions for over 100 variables

What can we see from space?

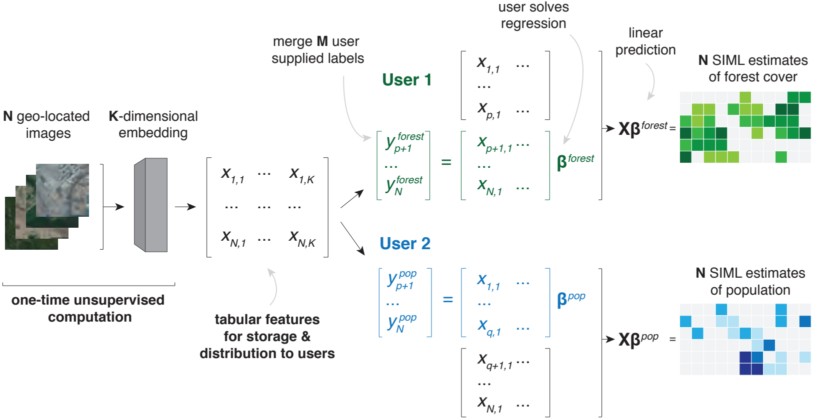
- Existing satellite imagery and machine learning (SIML) predictions are customized and heterogeneous
- This limits our ability to assess what can and cannot be feasibly monitored using this new technology

In this paper, we:

- **Assemble ground truth data** for 115 outcomes at national to global scales
- Build a flexible pipeline to **mass produce** MOSAIKS maps/predictions
- **Systematically examine** which categories of variables, geographic locations, and populations are most amenable to SIML-based monitoring
- **Distribute 115 maps** and associated data quality layers

Recall: MOSAIKS pipeline

Multi-task Observation using Satellite Imagery and Kitchen Sinks



Rolf, Proctor, Carleton, Bolliger, Shankar, Ishihara, Recht, & Hsiang (2021)

Pipeline extensions

Original MOSAIKS pipeline

7 tasks

Sparse sampling

By-hand label customization

Posted replication code

Pipeline extensions

Original MOSAIKS pipeline		This work
7 tasks	→	115 tasks
Sparse sampling	→	Dense sampling
By-hand label customization	→	Automated variable handling
Posted replication code	→	Interactive pipeline and resources

What are the limits of a planetary mapping pipeline?

For what types of variables does it work for?
What can we learn about the promise of satellite imagery?

Outline

Methods

Results

Summary

Outline

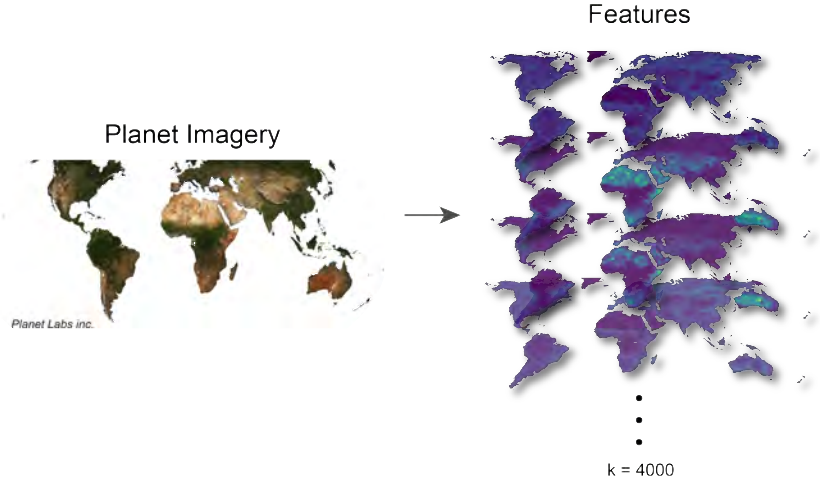
Methods

Results

Summary

One-time feature extraction

Random Convolutional Features (Rahimi & Recht (2007, 2008a,b))



details

Automated model specification

$$\mathbf{y}_l = X\boldsymbol{\beta}_l + \boldsymbol{\epsilon}_l$$

Automated model specification

$$y_l = X\boldsymbol{\beta}_l + \boldsymbol{\epsilon}_l$$

$$\downarrow$$

We automatically test variations of the ridge regression to find the optimal specification

Transformation: log or levels?

Intercept in regression: yes or no?

Use regional model: yes or no?

If polygon use weighting: area or population weighted?

Data gathering

We gather publicly available data with a diversity in spatial structures and categories.

Data gathering

We gather publicly available data with a diversity in spatial structures and categories.

Category	Number of Labels	Example Label
Agricultural Assets	5	Agricultural land ownership

Data gathering

We gather publicly available data with a diversity in spatial structures and categories.

Category	Number of Labels	Example Label
Agricultural Assets	5	Agricultural land ownership
Agriculture	16	Maize yield
Built Infrastructure	9	Buildings

Data gathering

We gather publicly available data with a diversity in spatial structures and categories.

Category	Number of Labels	Example Label
Agricultural Assets	5	Agricultural land ownership
Agriculture	16	Maize yield
Built Infrastructure	9	Buildings
Demographics	5	Median age
Education	10	Expected years of schooling
Health	15	Malaria in children
Household Assets	21	Mobile phones
Income	9	Human development index
Natural Systems	8	Tree cover
Occupation	17	Unemployment

Outline

Pipeline performance

Labels that...	have been remote sensed before	56
	have been remote sensed before with covariates	11
	have not been remote sensed before	48
	Total Labels	**115**

Pipeline performance

Labels that...	have been remote sensed before	56
	have been remote sensed before with covariates	11
	have not been remote sensed before	48
	Total Labels	**115**

Training time for one label

Run with autotuning: 20 CPU Hours

Pipeline performance

Labels that...	have been remote sensed before	56
	have been remote sensed before with covariates	11
	have not been remote sensed before	48
	Total Labels	**115**

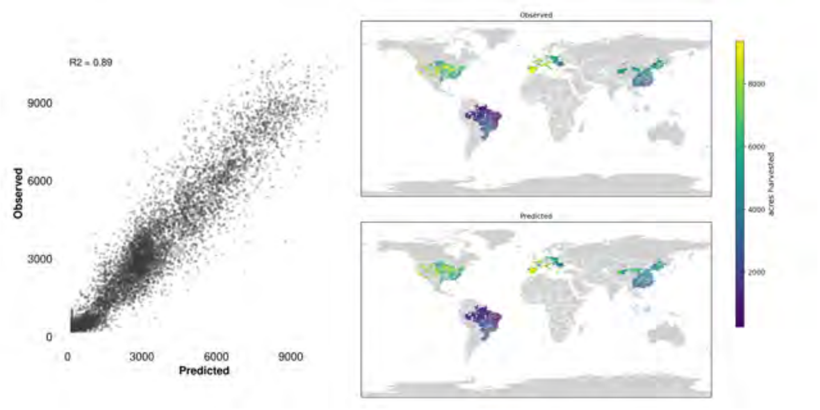
Training time for one label

Run with autotuning: 20 CPU Hours

Run final model: 2 minutes on laptop

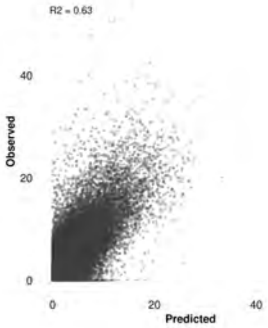
Maize yield

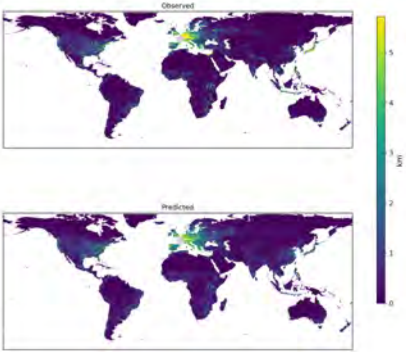
Source: Proctor (2021)
Description: Amount of maize harvested in acres



Road length

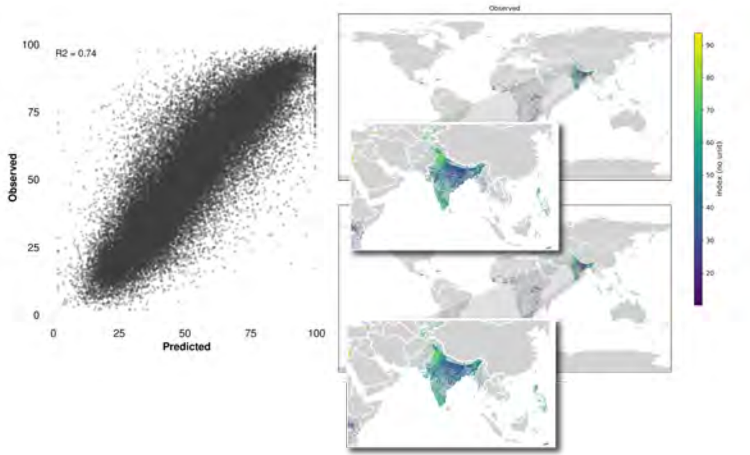
Source: Open Street Maps
Description: Length of roads in kilometers





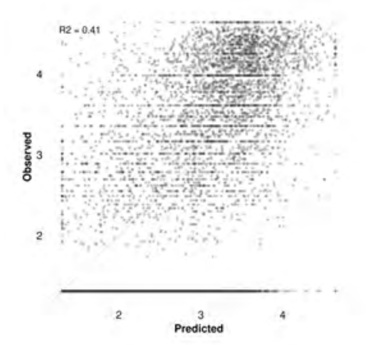
International wealth Index

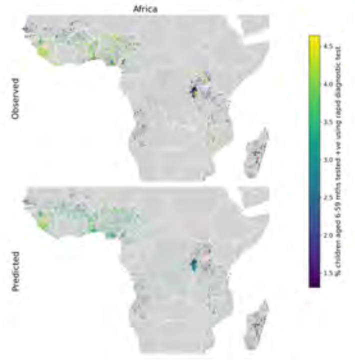
Source: Demographic and Health Surveys
Description: Mean international wealth index (IWI) for each DHS cluster



Malaria cases among children

Source: Demographic and Health Surveys
Description: Percentage of children age 6-59 months tested using a rapid diagnostic test (RDT) who are positive for malaria



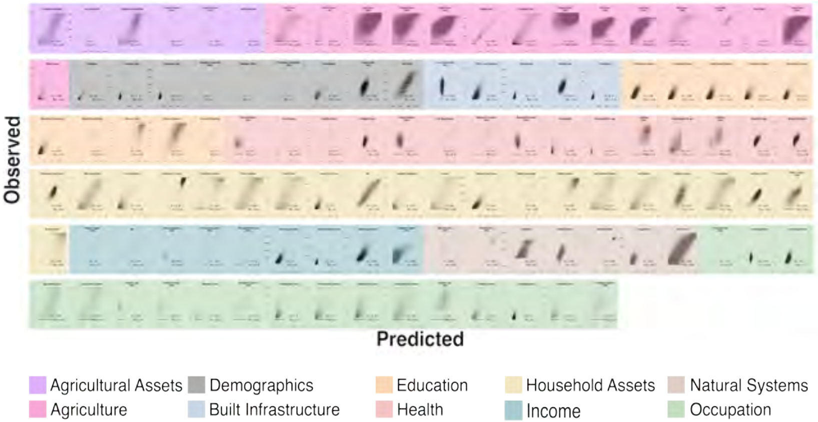


Petroleum pipelines

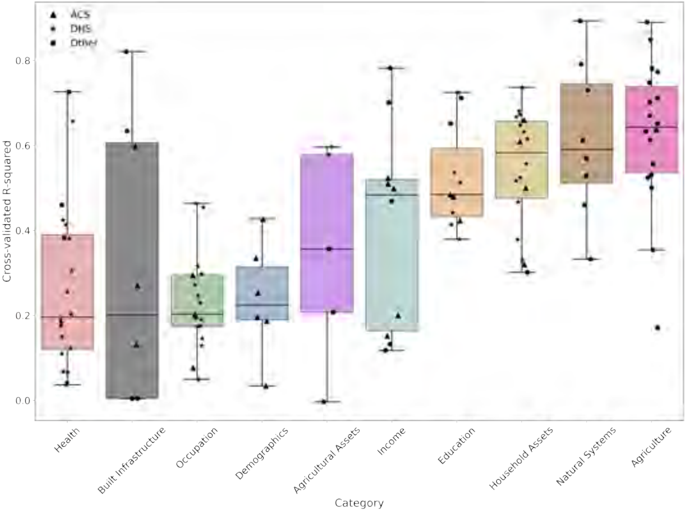
Source: Energy Information Administration
Description: Length of major petroleum product pipelines in kilometers

Pipelines not visible in imagery

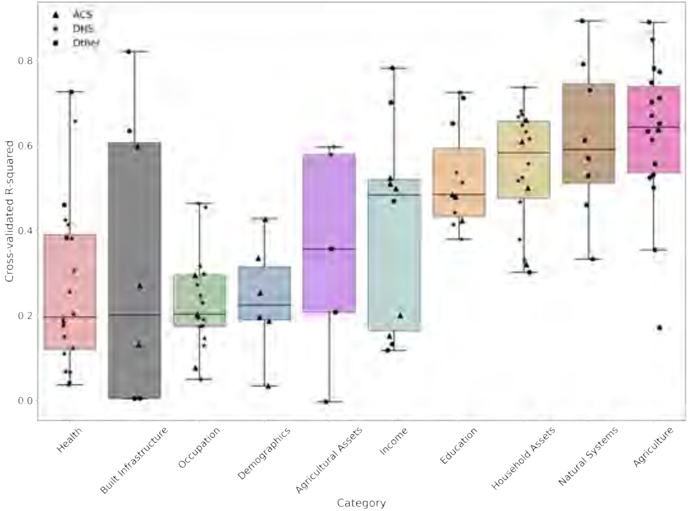
Performance across labels



Which label categories are amenable to SIML?

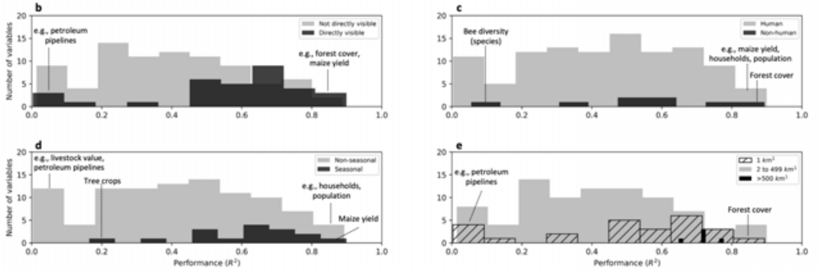


Which label categories are amenable to SIML?

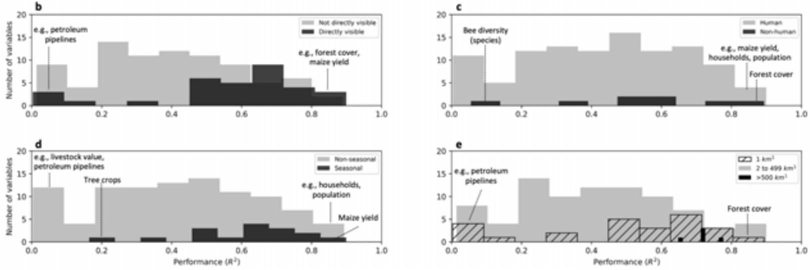


- Large variation in performance across, but especially <u>within</u> categories: σ_{within} = 0.19 and σ_{across} = 0.12

Which label categories are amenable to SIML?



Which label categories are amenable to SIML?



- Directly visible versus emergent: difference in mean R^2 = 0.18 ($p < 0.001$)
- Clear seasonal cycle versus without: difference in mean R^2 = 0.19 ($p < 0.05$)
- Human versus natural systems: difference in mean R^2 = 0.1 ($p = 0.23$)
- Spatial resolution of ground-truth data: all differences in mean $p > 0.1$.

Where are SIML measurements reliable?

Mean Absolute Normalized Error

$$\text{MANE}_i = \frac{1}{|L_i|} \sum_{l \in L_i} \left| \frac{\hat{\varepsilon}_{li} - \overline{\hat{\varepsilon}}_l}{\sigma(\hat{\varepsilon}_{li})} \right|$$

- $\hat{\varepsilon}_{li}$: residual of pixel i of label l
- $\overline{\hat{\varepsilon}}_l$: mean of residuals for label l
- $\sigma(\hat{\varepsilon}_{li})$: standard deviation for residuals label l in pixel i
- L_i: the set of labels we have in pixel i

Where are SIML measurements reliable?

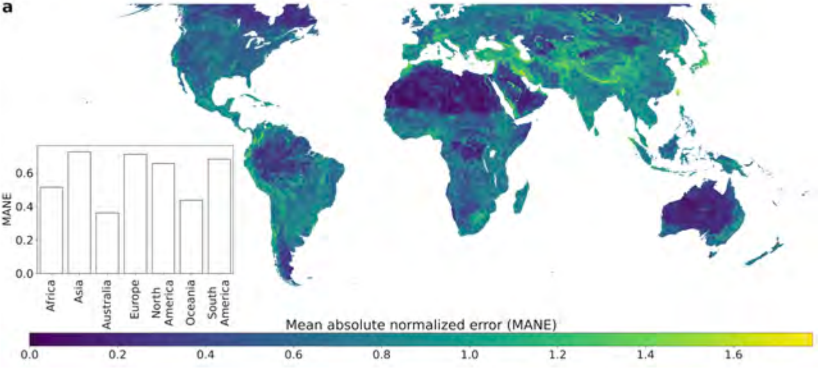
Mean Absolute Normalized Error

$$\text{MANE}_i = \frac{1}{|L_i|} \sum_{l \in L_i} \left| \frac{\hat{\varepsilon}_{li} - \overline{\hat{\varepsilon}}_l}{\sigma(\hat{\varepsilon}_{li})} \right|$$

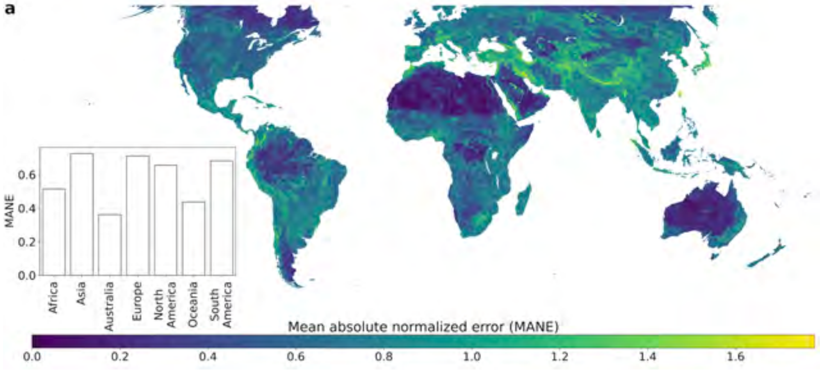
- $\hat{\varepsilon}_{li}$: residual of pixel i of label l
- $\overline{\hat{\varepsilon}}_l$: mean of residuals for label l
- $\sigma(\hat{\varepsilon}_{li})$: standard deviation for residuals label l in pixel i
- L_i: the set of labels we have in pixel i

MANE $\approx$ z-score

Where are SIML measurements reliable?



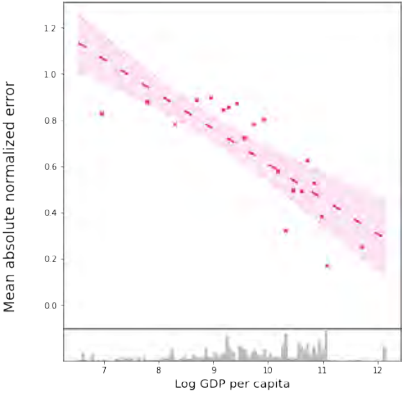
Where are SIML measurements reliable?



- MANE differs significantly across continents: $p < 0.01$
- Lowest errors in Australia and Africa (deserts) and highest errors in North America and Asia (urban areas and high elevation)
- High errors in extreme locations $\longrightarrow$ mean-reverting measurement error (common in ML)

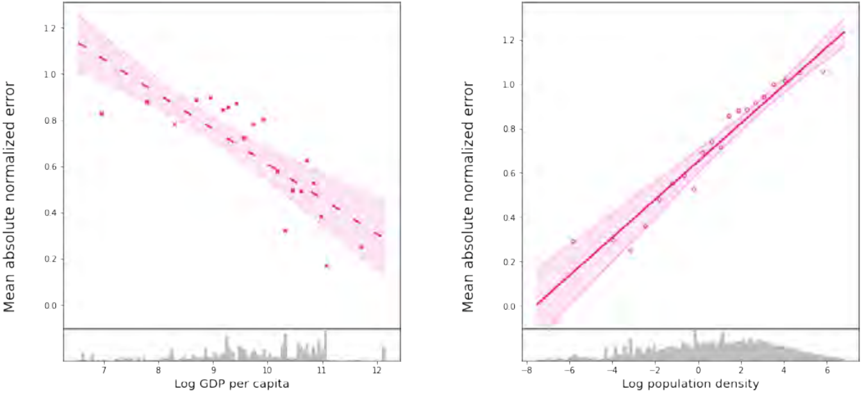
Where are SIML measurements reliable?

Example: Agriculture



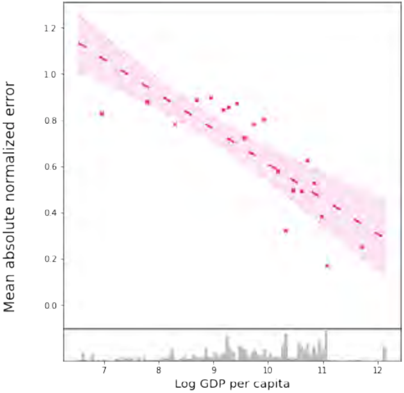
Where are SIML measurements reliable?

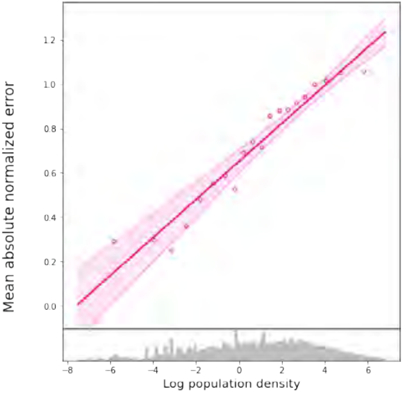
Example: Agriculture



Where are SIML measurements reliable?

Example: Agriculture

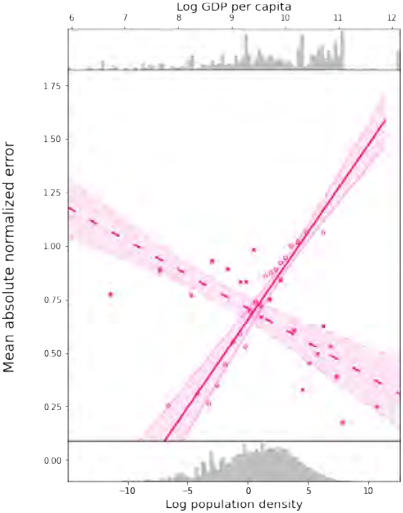




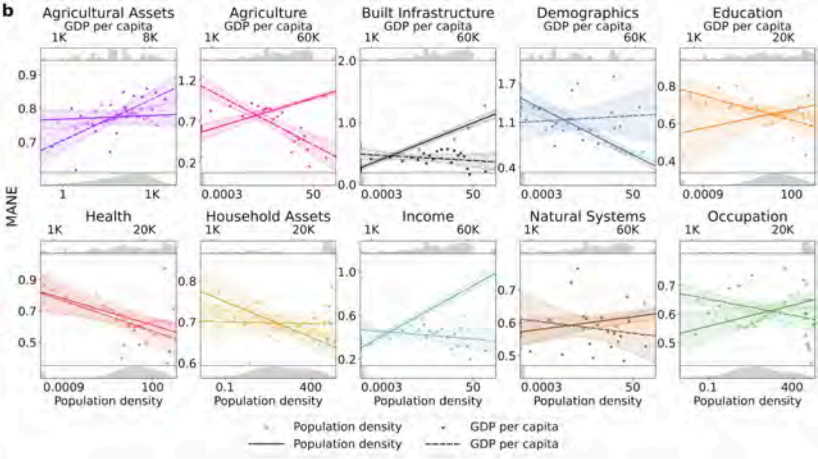
Put it together...

Where are SIML measurements reliable?

Example: Agriculture

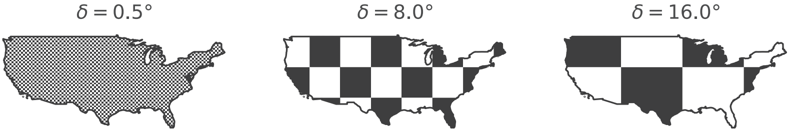


Where are SIML measurements reliable?



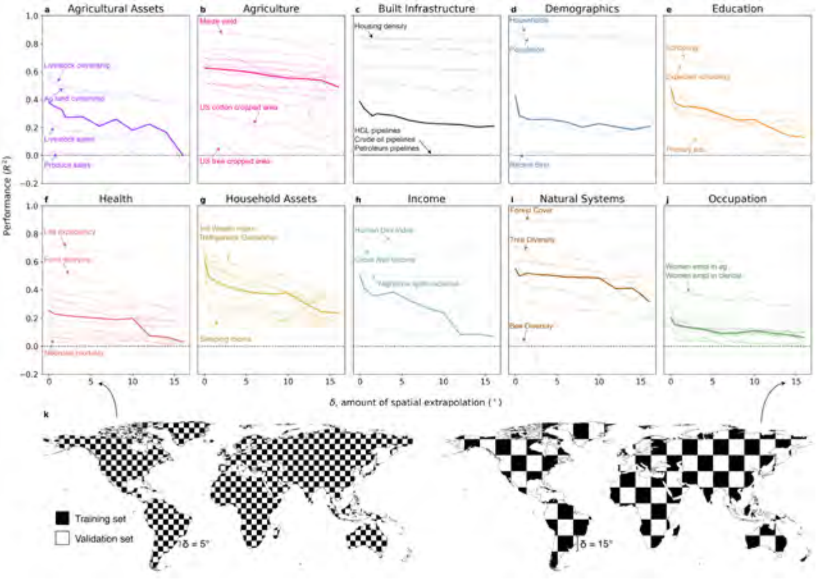
How far from training data can we extrapolate?

1. Partition sample in checkerboard
2. Train on white squares
3. Test on black squares
4. Jitter checkerboard location & repeat
5. Compare to spatial interpolation of ground-truth



(Reference: 8° $\times$ 8° = 888 km $\times$ 682 km (552 mi $\times$ 424 mi) at centroid)

How far from training data can we extrapolate?



Assessing and communicating uncertainty

We use the results from spatial cross-validation to report estimated data quality alongside release of global predictions

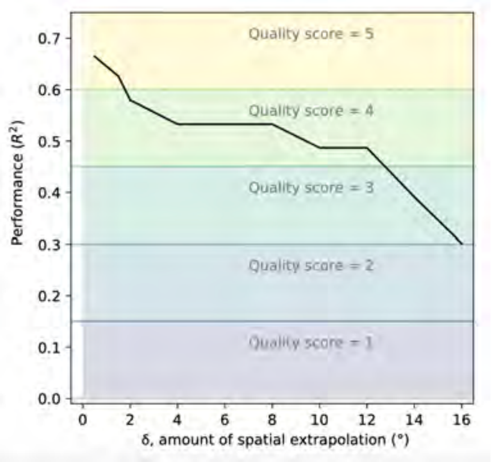
For each of 115 outcomes:

1. Conduct spatial cross-validation (i.e., “checkerboard”) experiment detailed above
2. Determine estimated R^2 for all locations outside training set based on distance to nearest ground truth observation
3. Assign all locations a quality score based on estimated R^2: 0 ($R^2 < 0$) $\rightarrow$ 5 ($R^2 \geq 0.6$)
4. For socioeconomic outcomes, mask populated areas

Assessing and communicating uncertainty

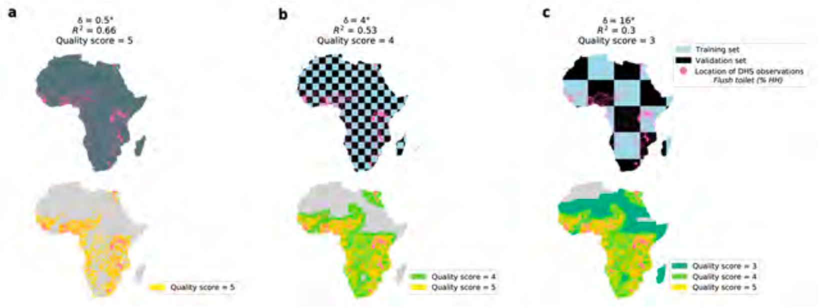
1. Conduct spatial cross-validation (i.e., “checkerboard”) experiment detailed above

E.g., access to flush toilets:

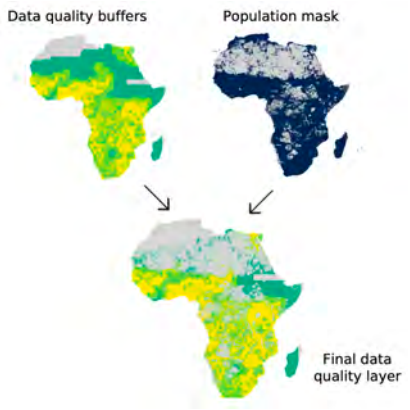


Assessing and communicating uncertainty

2. Determine estimated R^2 for all locations outside training set based on distance to nearest ground truth observation
3. Assign all locations a quality score based on estimated R^2: 0 ($R^2 < 0$) $\rightarrow$ 5 ($R^2 \geq 0.6$)

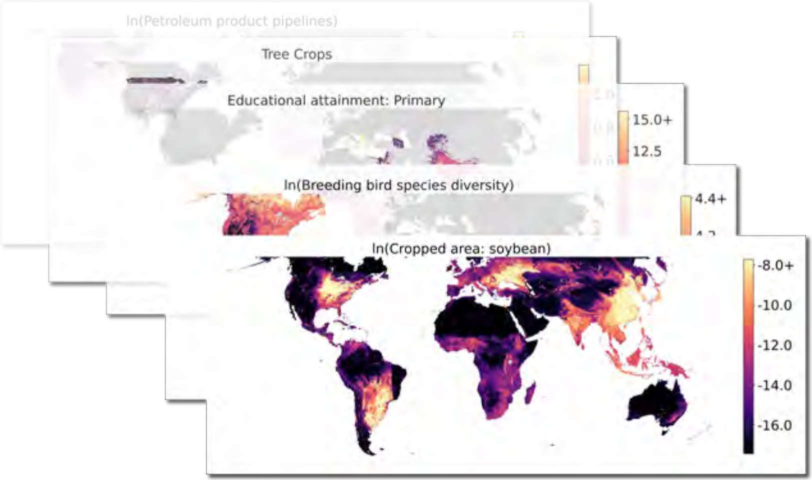


4. For socioeconomic outcomes, mask populated areas

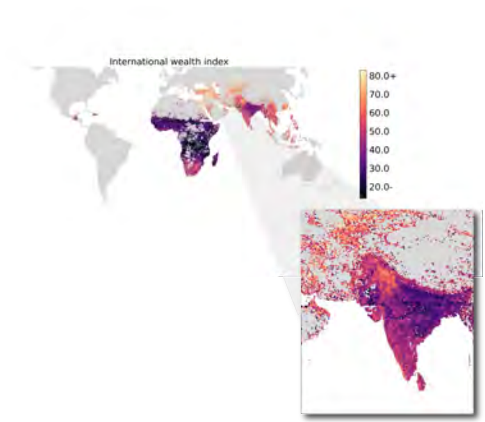


Distribute predictions and data quality layers

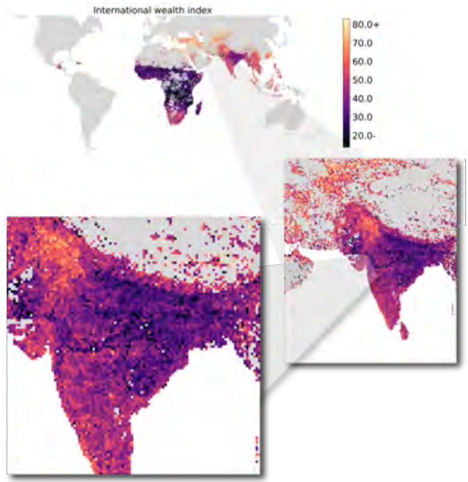
115 Maps



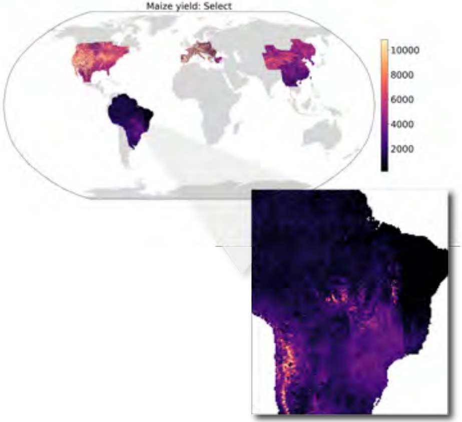
Example map layers



Example map layers

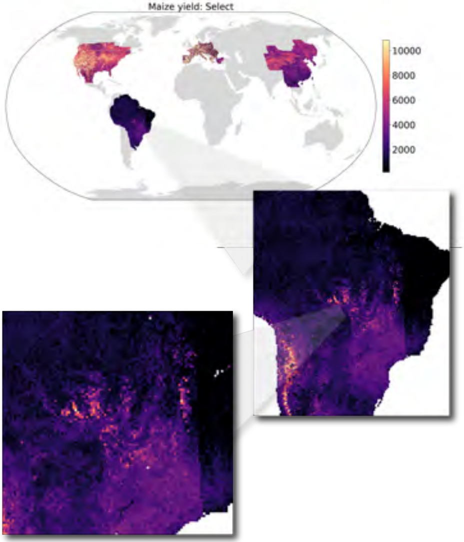


Example map layers

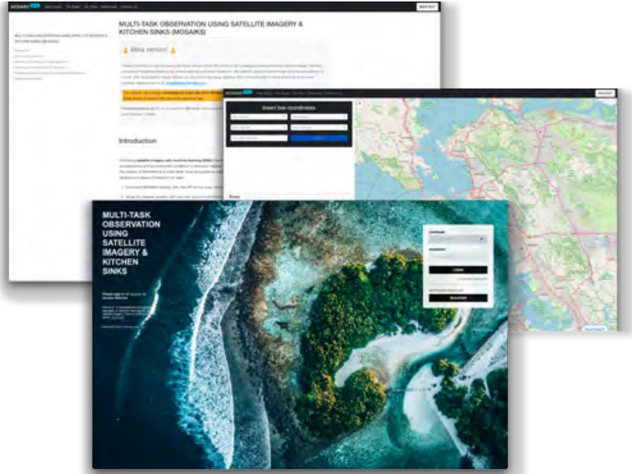


details

Example map layers



MOSAIKS API - `www.mosaiks.org`



Outline

Methods

Results

Summary

Conclusions

1. We demonstrate that mass production of global maps from satellite imagery is possible
   - $>$100 maps predicted using a single featurization of imagery, at low computational cost
   - Automated tuning pipeline makes imagery-based predictions “plug-and-play”

Conclusions

1. **We demonstrate that mass production of global maps from satellite imagery is possible**
    - >100 maps predicted using a single featurization of imagery, at low computational cost
    - Automated tuning pipeline makes imagery-based predictions "plug-and-play"
2. **We uncover patterns that characterize the promise of satellite imagery for global mapping**
    - Observe high (agriculture, household assets, natural systems) and low (health, demographics) performers
    - Show how prediction error correlates with location, income, population

Conclusions

1. We demonstrate that mass production of global maps from satellite imagery is possible
    - $>$100 maps predicted using a single featurization of imagery, at low computational cost
    - Automated tuning pipeline makes imagery-based predictions "plug-and-play"
2. We uncover patterns that characterize the promise of satellite imagery for global mapping
    - Observe high (agriculture, household assets, natural systems) and low (health, demographics) performers
    - Show how prediction error correlates with location, income, population
3. We publicly release inputs, outputs, and training resources
    - API: 4,000 features from Planet Labs, Inc., extensive tutorial resources
    - Data quality layer provided alongside predicted maps
    - All code and input data public


MOSAIKS

Acknowledgements

Team
Benjamin Recht, Esther Rolf, Eugenio Noda, Hikari Murayama, Hannah Druckenmiller, Ian Bolliger, Jeanette Tseng, Jessica Katz, Jonathan Proctor, Luke Sherman, Miyabi Ishihara, Simon Greenhill, Solomon Hsiang, Tamma Carleton, Taryn Fransen, Trinetta Chong, Vaishaal Shankar

Partners and Funding
USAID, UNDP, CEGA, UCSB Bren School: Data Science students

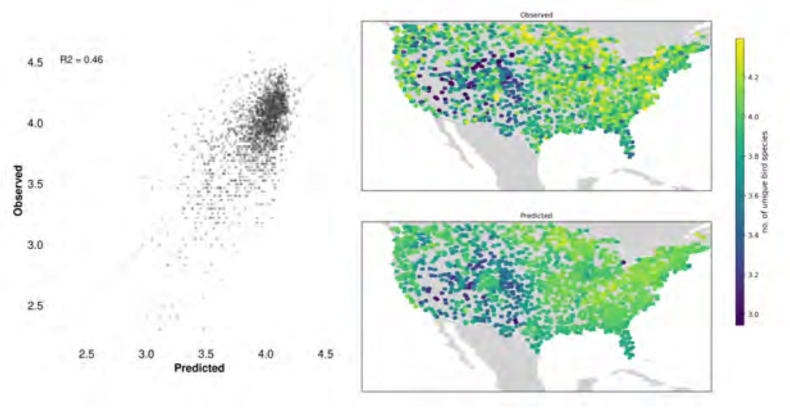
MOSAIKS API

www.mosaiks.org

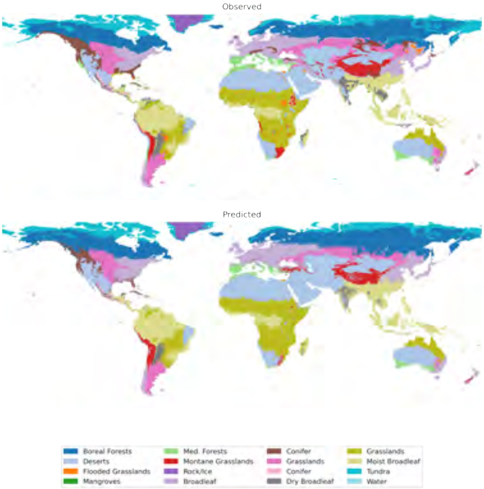
Appendix

Breeding bird species diversity

Source: U.S. Geological Survey
Description: Count of the unique species documented by all road segments



Classifier Example



Test and cross-validated performance